



**caBIG<sup>®</sup>**

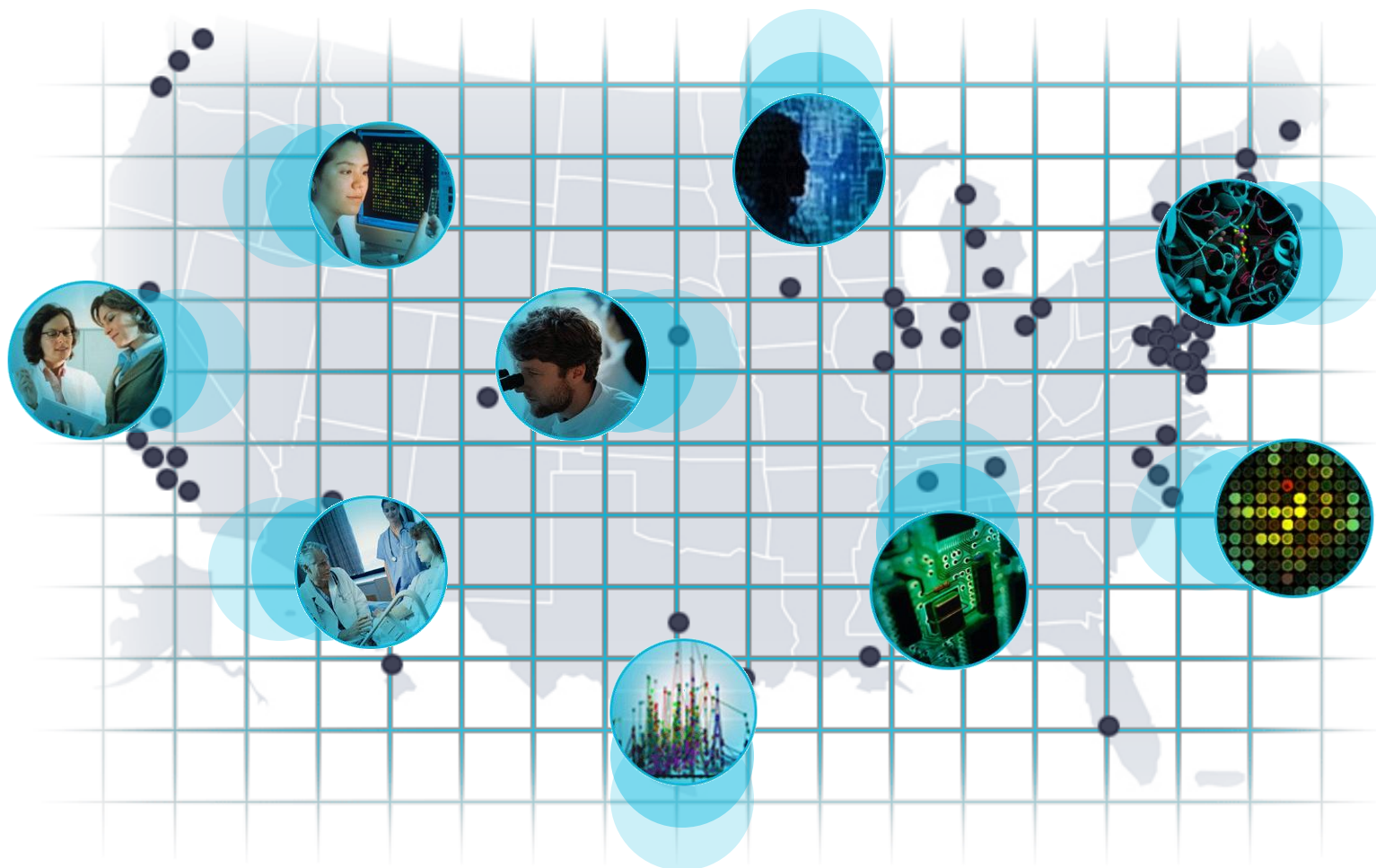
cancer Biomedical  
Informatics Grid<sup>™</sup>

# **The cancer Biomedical Informatics Grid (caBIG<sup>®</sup>)**

Kenneth H. Buetow, Ph.D.  
National Cancer Institute



# NCI is Utilizing Informatics to Integrate Cancer Information



# The caBIG<sup>®</sup> Initiative



caBIG<sup>®</sup> is an a virtual web of interconnected data, individuals, and organizations that redefines how research is conducted, care is provided, and patients/participants interact with the biomedical research enterprise.

## caBIG<sup>®</sup> Vision

- **Connect** the cancer research community through a shareable, interoperable infrastructure
- **Deploy and extend** standard rules and a common language to more easily share information
- **Build or adapt** tools for collecting, analyzing, integrating and disseminating information associated with cancer research and care

# caBIG<sup>®</sup> Capabilities Enable

Discovery > Clinical Research > Clinical Care



## Molecular Medicine



**Clinical Research**



**Imaging**



**Molecular Biology**



**Pathology**



# caBIG<sup>®</sup> Capabilities Enable

## Discovery > Clinical Research > Clinical Care



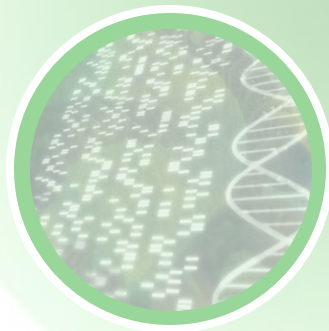
### Molecular Medicine

- Track clinical trial registrations
- Facilitate automatic capture of clinical laboratory data
- Manage reports describing adverse events during clinical trials



Clinical Research

- Combine proteomics, gene expression, and other basic research data
- Submit and annotate microarray data
- Integrate microarray data from multiple manufacturers and permit analysis and visualization of data



Molecular Biology



Imaging

- Utilize the National Cancer Imaging Archive repository for medical images including CAT scans and MRIs
- Visualize images using DICOM-compliant tools
- Annotated Images with distributed tools



Pathology

- Access a library of well characterized, clinically annotated biospecimens
- Use tools to keep an inventory of a user's own samples
- Track the storage, distribution, and quality assurance of specimens

# caBIG<sup>®</sup> and Clinical Research



## Sample capabilities and tools:

- **caBIG<sup>®</sup>-compatible Commercial Electronic Data Capture system available through Cancer Community-wide license (announcement of selection in next couple of weeks)**
- **Clinical data exchange (caXchange)**
- **Virtual clinical data warehouse (CTODS)**
- **Study participant registry (C3PR)**
- **Study participant calendar (PSC)**
- **Adverse event management (caAERS)**
- **caBIG<sup>®</sup>-compatible systems architecture (caGrid)**





## Sample capabilities and tools:

- **Biobanking management systems** (caTissue Core)
- **Microarray data management system** (caArray)
- **Genome-wide data management system** (caGWAS)
- ***In vivo* image repository** (NCIA)
- **Virtual clinical data warehouse** (CTODS)
- **Microarray gene expression and sequence data analysis** (geWorkbench)
- **caBIG<sup>®</sup> -compatible systems architecture** (caGrid)



# Data Sharing and Security



## Sample resources:

- **caBIG® Policies**
- **Processes and Best Practices**
- **Model Documents**







**caBIG<sup>®</sup>**

cancer Biomedical  
Informatics Grid<sup>™</sup>

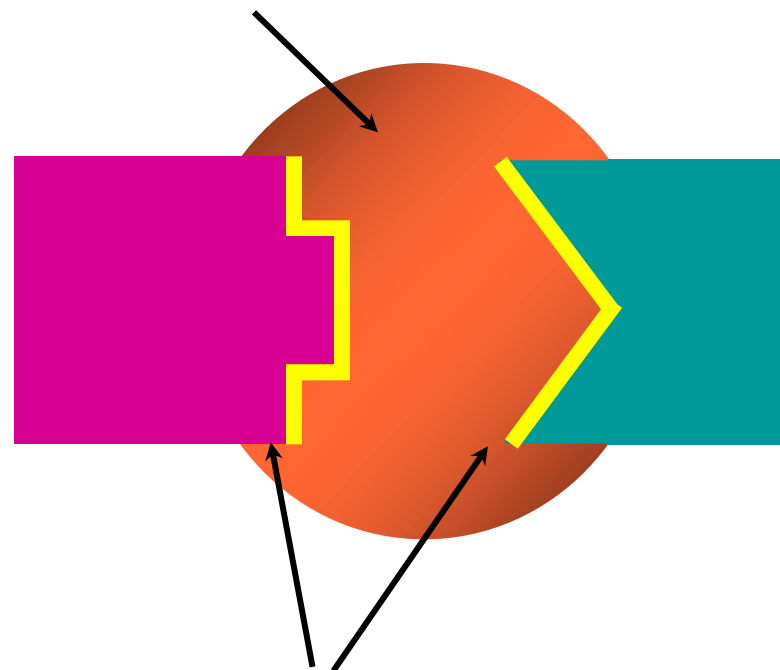
**caBIG<sup>®</sup>: an open  
SOA with  
shared  
community  
semantics**

# Boundaries and Interfaces



- **Focus on boundaries and interfaces, how things fit together, not on the internal details**
- **Once they're built: assume that will be diverse & changing**

The glue that binds parts together is middleware infrastructure

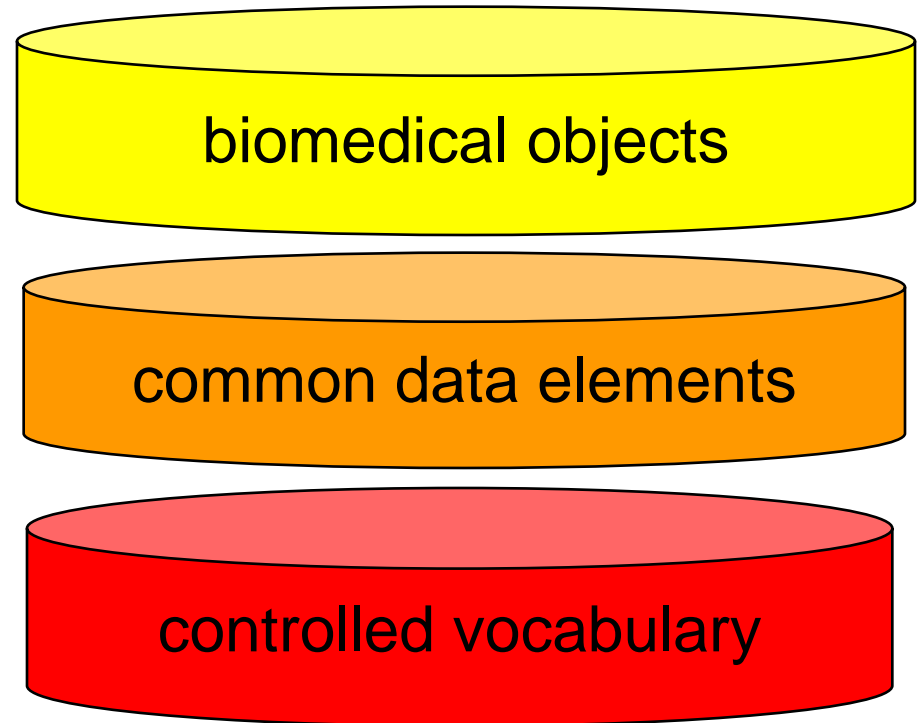


Shape of boundary is defined in APIs

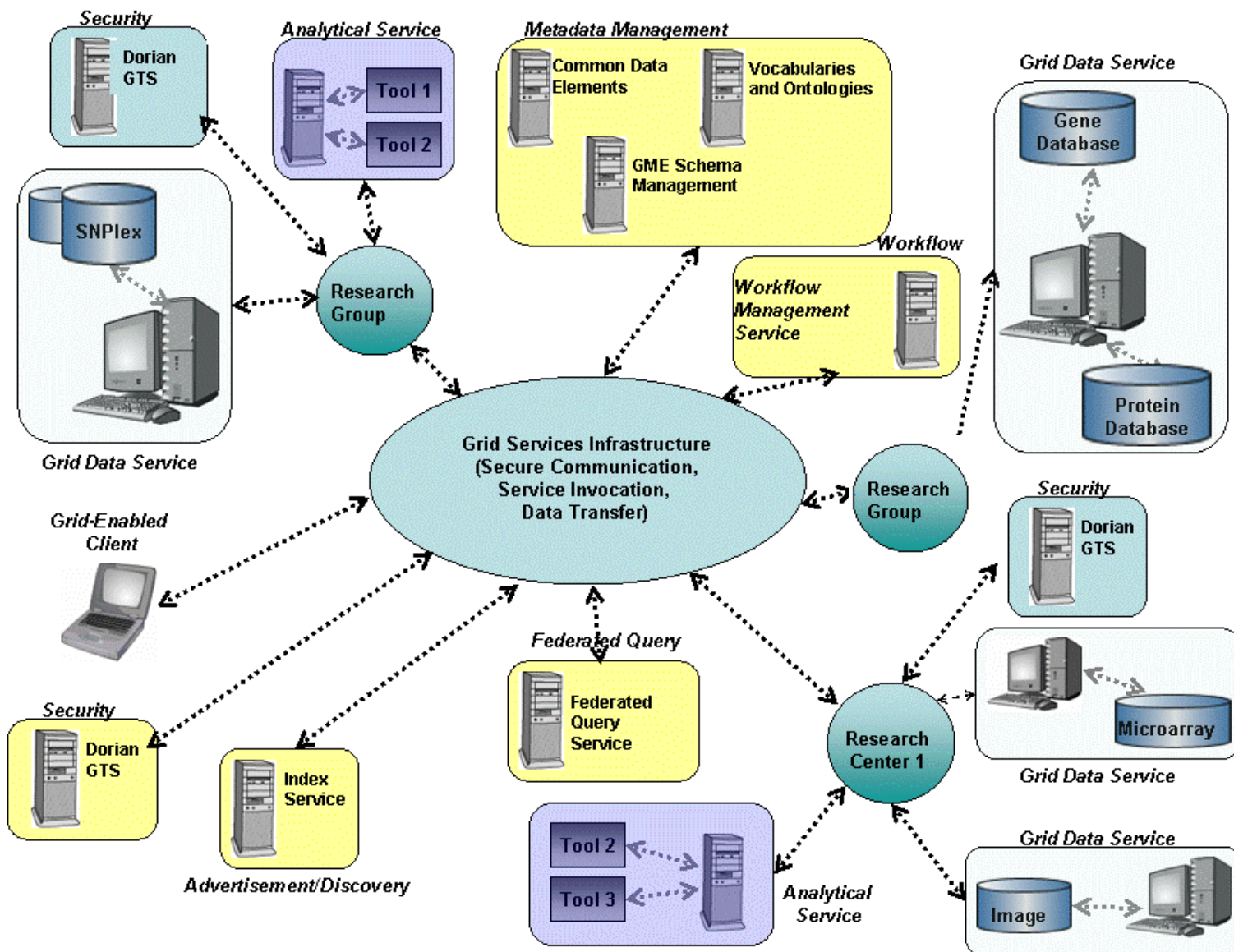
# Standards-based Interoperability: caCORE



- **Community driven**
- **Dynamic implementation**
  - Built to be upgraded as standards “harden”, and domains expand



# caGrid 1.1 Conceptual View



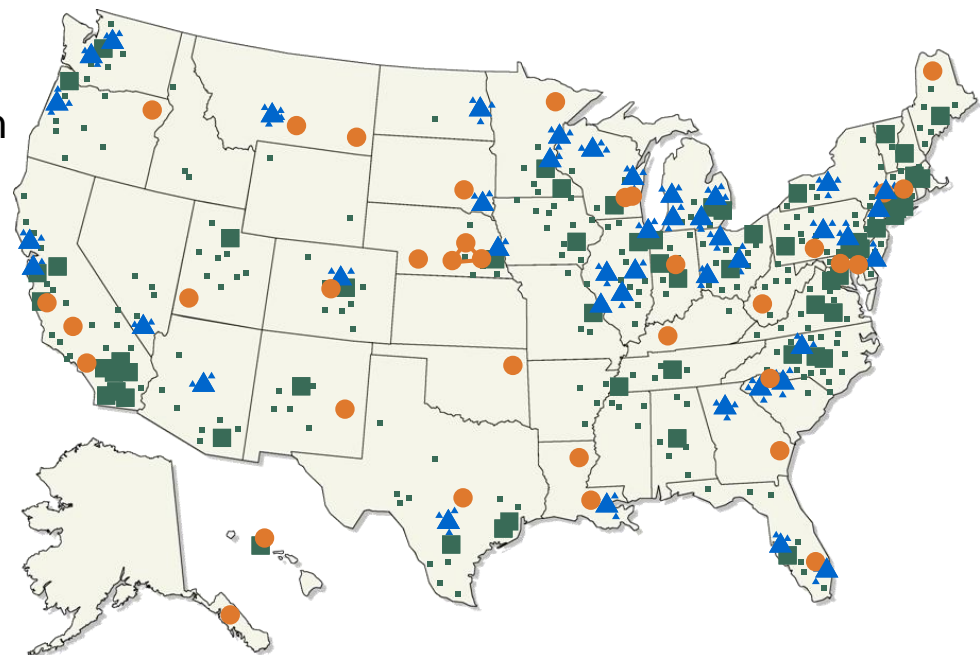


# Connected with caBIG<sup>®</sup>



- **caBIG<sup>®</sup> adoption is unfolding in:**
  - 56 NCI-designated Cancer Centers
  - 16 NCI Community Cancer Centers
- caBIG<sup>®</sup> being integrated into federal health architecture to connect **National Health Information Network**
- **Global Expansion**
  - United Kingdom
  - China
  - India
  - Latin America

## NCI-Designated Cancer Centers, Community Cancer Centers, and Community Oncology Programs



# caBIG<sup>®</sup> Enterprise Support Network:

## *Knowledge Centers*



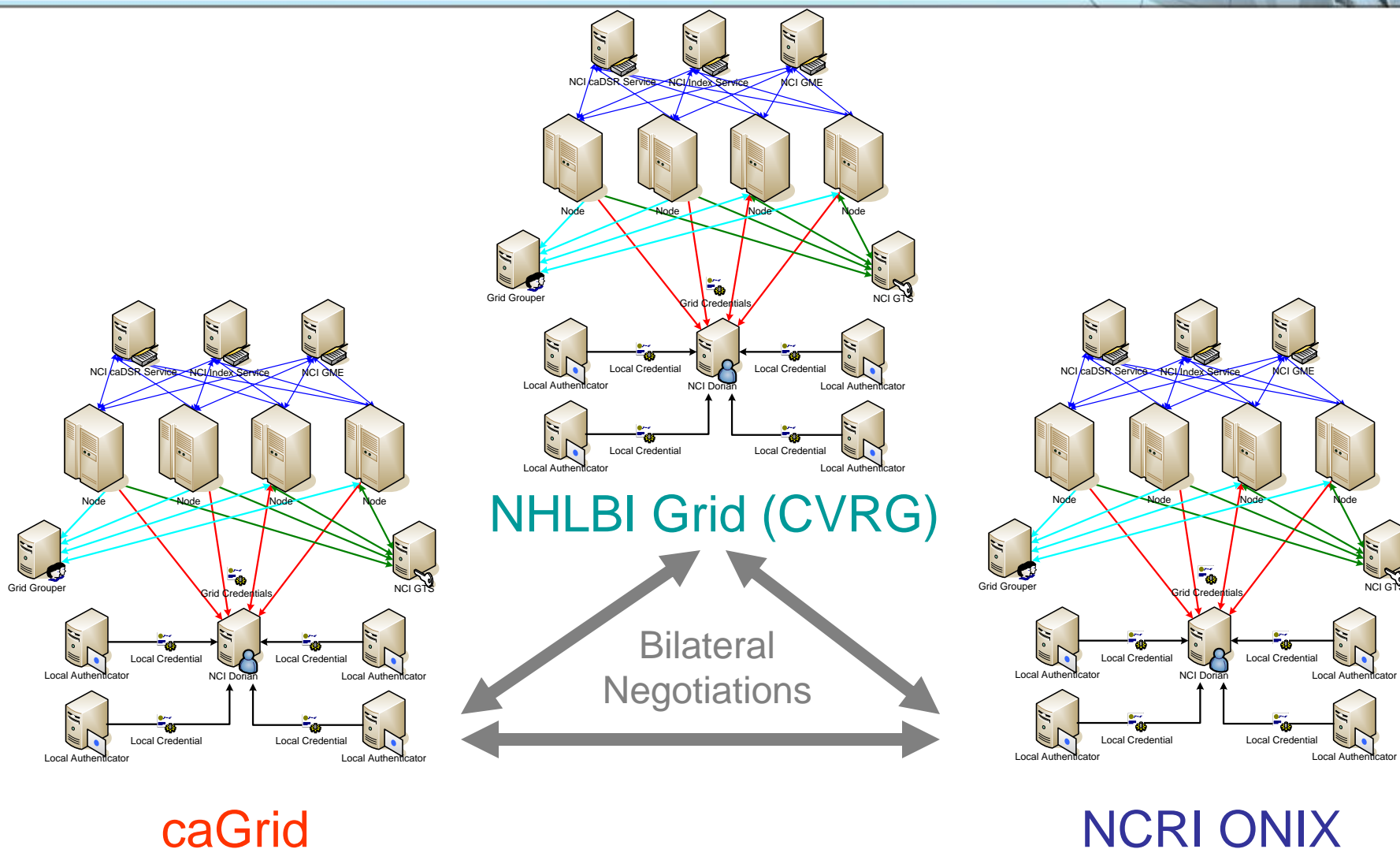
- **caGrid:** The Ohio State University and The Ohio State Comprehensive Cancer Center, with the University of Chicago and the Argonne National Lab
- **Clinical Trials Management Systems:** Duke University Comprehensive Cancer Center, with Robert H. Lurie Comprehensive Cancer Center at Northwestern University, Cancer and Leukemia Group B – Information Systems (CALGB-IS), and SemanticBits
- **Data Sharing and Intellectual Capital:** University of Michigan
- **Molecular Analysis Tools:** Columbia University Herbert Irving Comprehensive Cancer Center with The Broad Institute of MIT and Harvard
- **Tissue/Biospecimen Banking and Technology Tools:** Siteman Cancer Center, Washington University at St. Louis
- **Vocabulary:** Mayo Clinic with SemanticBits

# Interoperability Goes Global

- **UK National Cancer Research Institute** (Shared interoperable biomedical infrastructure)
- **China/Shanghai Center for Bioinformatics Technology Beijing Cancer Hospital** (caBIG® technology for life sciences and clinical research)
- **Indian National Knowledge Commission** (Nationwide IT infrastructure linking all universities, major hospitals, research labs and libraries; cancer and beyond)



# Grid of Grids...







**caBIG<sup>®</sup>**

cancer Biomedical  
Informatics Grid<sup>™</sup>

# **caBIG<sup>®</sup> and The Cancer Genome Atlas (TCGA)**

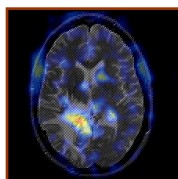




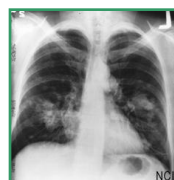
## Connecting multiple sources, experiments, and data types

### Three forms of cancer

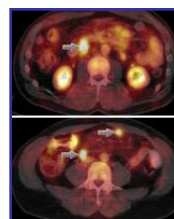
glioblastoma multiforme  
(brain)



squamous carcinoma  
(lung)



serous  
cystadenocarcinoma  
(ovarian)



### 12 Organizations

Biospecimen Core  
Resource

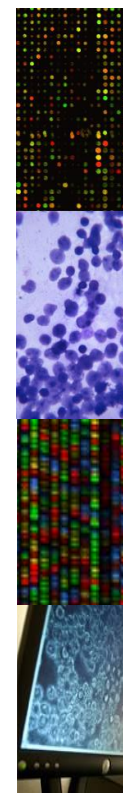
7 Cancer Genomic  
Characterization  
Centers

3 Genome  
Sequencing  
Centers

Data Coordinating  
Center

### Multiple data types

- Clinical diagnosis
- Treatment history
- Histologic diagnosis
- Pathologic status
- Tissue anatomic site
- Surgical history
- Gene expression
- Chromosomal copy number
- Loss of heterozygosity
- Methylation patterns
- miRNA expression
- DNA sequence



A single web-based portal for all analyses – <http://cma.nci.nih.gov>



National Cancer Institute

Context: TCGA ▼



**caBIG**<sup>TM</sup> cancer Biomedical  
Informatics Grid<sup>TM</sup>

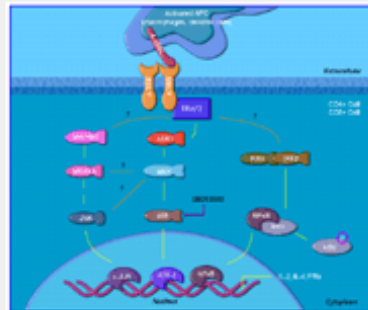
Cancer Molecular Analysis Portal

## Gene View

### Gene View

Visualize gene expression, copy number, SNP, and pathway data on a gene by gene basis. Generate detailed study related reports for a given gene.

Available resources include: Gene Expression Plots, KM Survival Plots, CGWB Integration, and Pathway Visualizations.



## Genome View

## Clinical View

## Analysis Tools

### Existing Users:

user:

pass:

### Additional Information:

- ♦ Register
- ♦ Provide your feedback



FIRSTGOV

A single web-based portal for all analyses – <http://cma.nci.nih.gov>



National Cancer Institute

Context: TCGA ▼



**caBIG**<sup>TM</sup> cancer Biomedical  
Informatics Grid<sup>TM</sup>

Cancer Molecular Analysis Portal

Gene View

**Genome View**

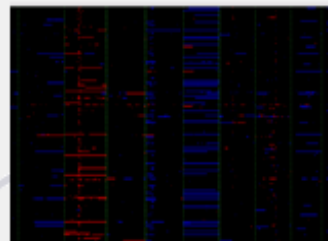
Clinical View

Analysis Tools

### Genome View

Explore all of the study data in one genome level visualization. Investigate chromosomal regions of amplification, deletion and over expression. Zoom in on a chromosomal region of interest for a more detailed view.

Available resources include: Integrated Heatmap Viewer for Genomic Data.



### Existing Users:

user:

pass:

### Additional Information:

- ♦ [Register](#)
- ♦ [Provide your feedback](#)



FIRSTGOV



A single web-based portal for all analyses – <http://cma.nci.nih.gov>



# caBIG<sup>TM</sup> cancer Biomedical Informatics Grid<sup>TM</sup>

## Cancer Molecular Analysis Portal

Gene View

Genome View

Clinical View

Analysis Tools

### Clinical View

Investigate the study clinical data. Explore the relationships between clinical and molecular study data.

Available resources include: Clinical Query with Reports and KM Sample Plots



### Existing Users:

user:   
pass:

### Additional Information:

- ♦ Register
- ♦ Provide your feedback

A single web-based portal for all analyses – <http://cma.nci.nih.gov>



National Cancer Institute

Context: TCGA ▼



**caBIG™** cancer Biomedical  
Informatics Grid™

Cancer Molecular Analysis Portal

Gene View

Genome View

Clinical View

**Analysis Tools**

### Analysis Tools

Analyze the study data using analysis tools such as Genepattern, Principal Component Analysis, and the Cancer Genome Workbench

Available resources include: PCA Analysis and Gene Pattern Integration



**GenePattern**

### Existing Users:

user:

pass:

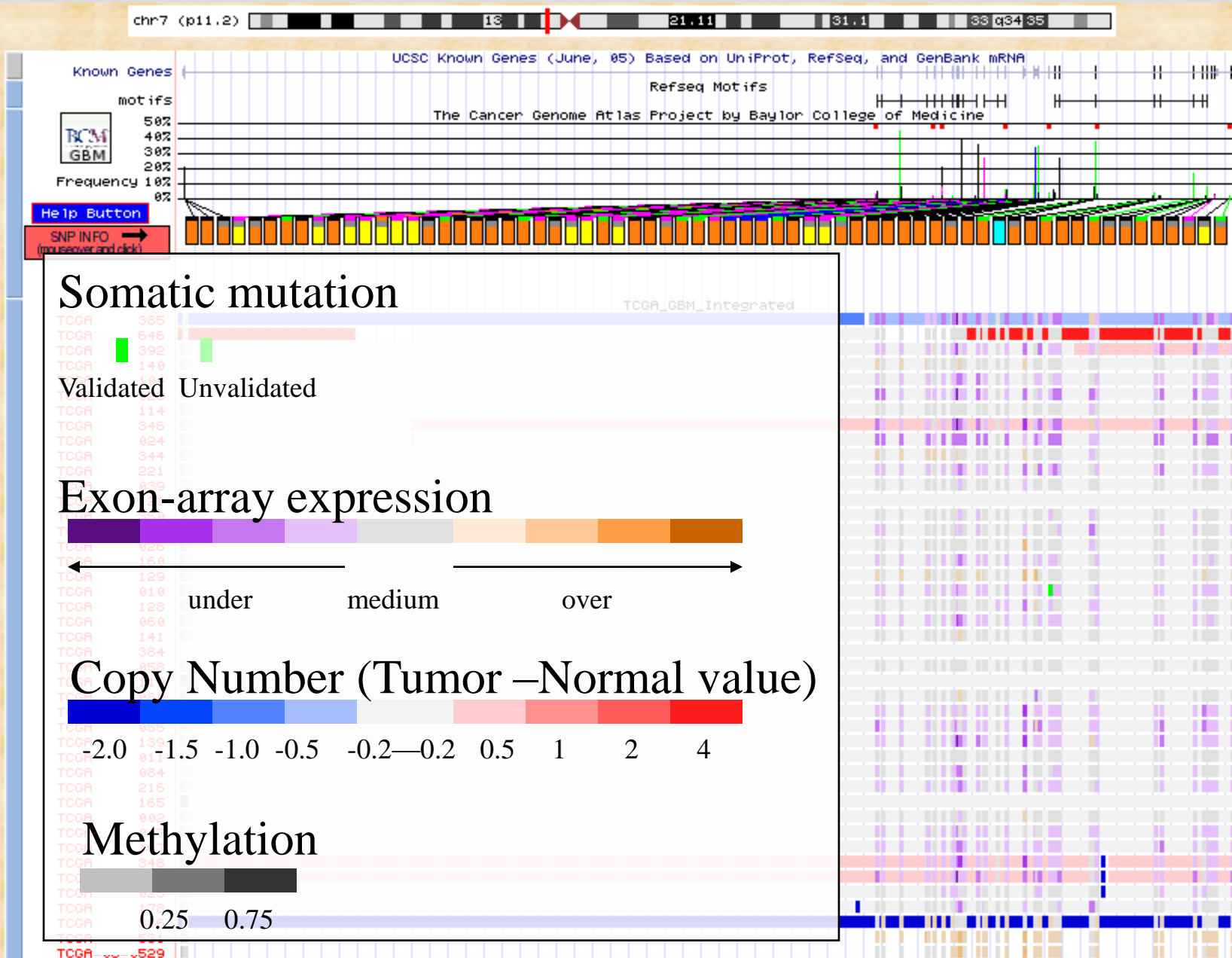
### Additional Information:

- ♦ [Register](#)
- ♦ [Provide your feedback](#)



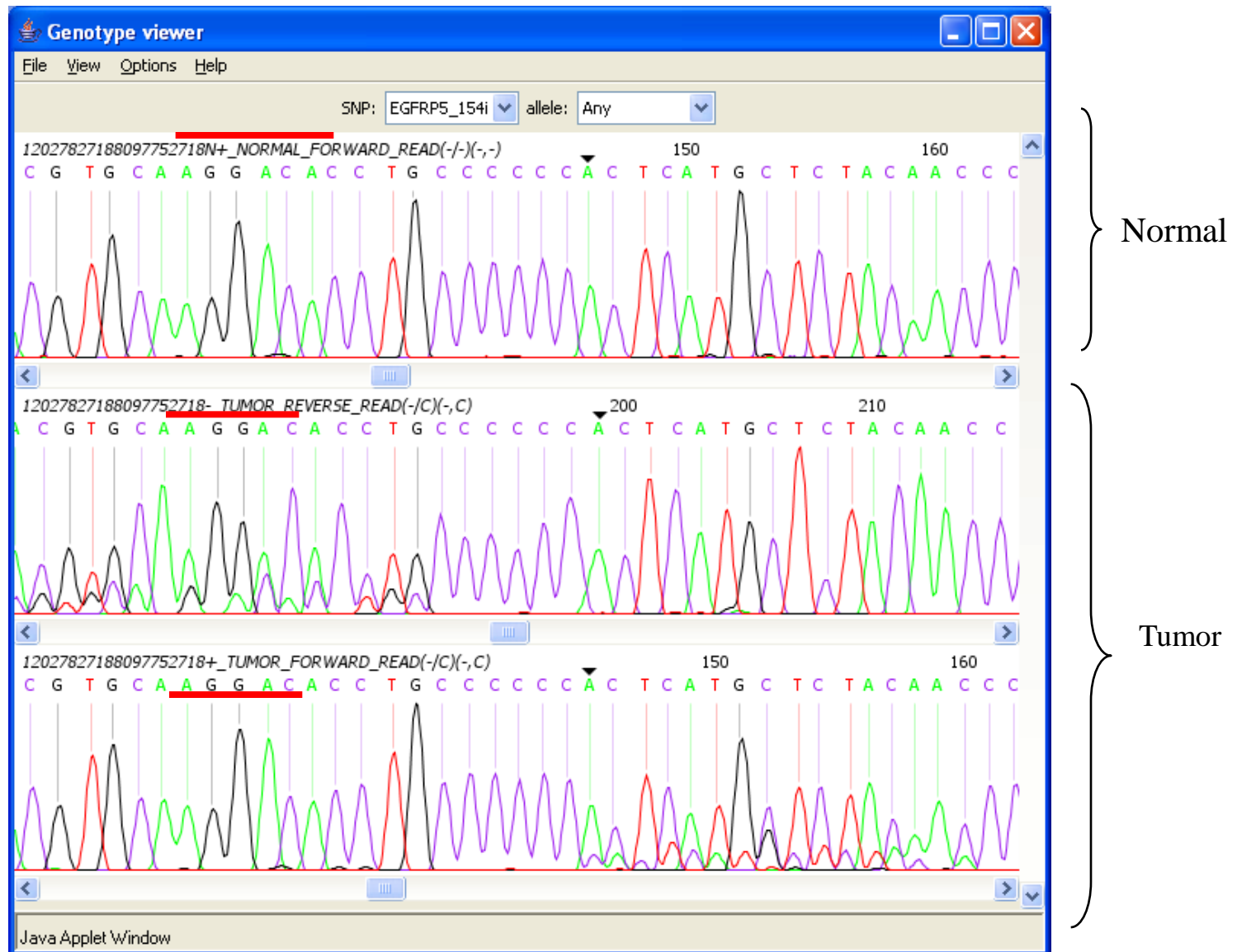
FIRSTGOV

# Comprehensive Summary



# Putative somatic mutations can be manually reviewed

## Example: a frameshift mutation in EGFR in paired tumor/normal





# Protein structure view of EGFR mutations

Protein Structure 3D viewer - Mozilla Firefox

File Edit View History Bookmarks Tools Help

https://cgwb.nci.nih.gov/cgi-bin/3dViewer/Gene.cgi?proj=valid&sym=EGFR

Human chr7:55,053,218-55,243,524 - ... Protein Structure 3D viewer

National Cancer Institute  
Protein Structure Viewer

Access proteins from alternative splicing

Gene : EGFR GI : 29725609 1210 aa Also see GI : 41327736 41327732 41327734

Protein Motifs : 4 Solid : Pfam domain Pdb hits: 1 cyan 3D Structure Viewing region(s): 25 to 638

SNP LogE & SIFT: Red : Predicted Deleterious Blue : Predicted Tolerant Black : Undecided Gray : Not Analyzed

Phosphorylate : 29 Pubmed Tabular format

Click to get 3D viewer

1 1000 1200

Done cgwb.nci.nih.gov

# 3D Structure Viewer

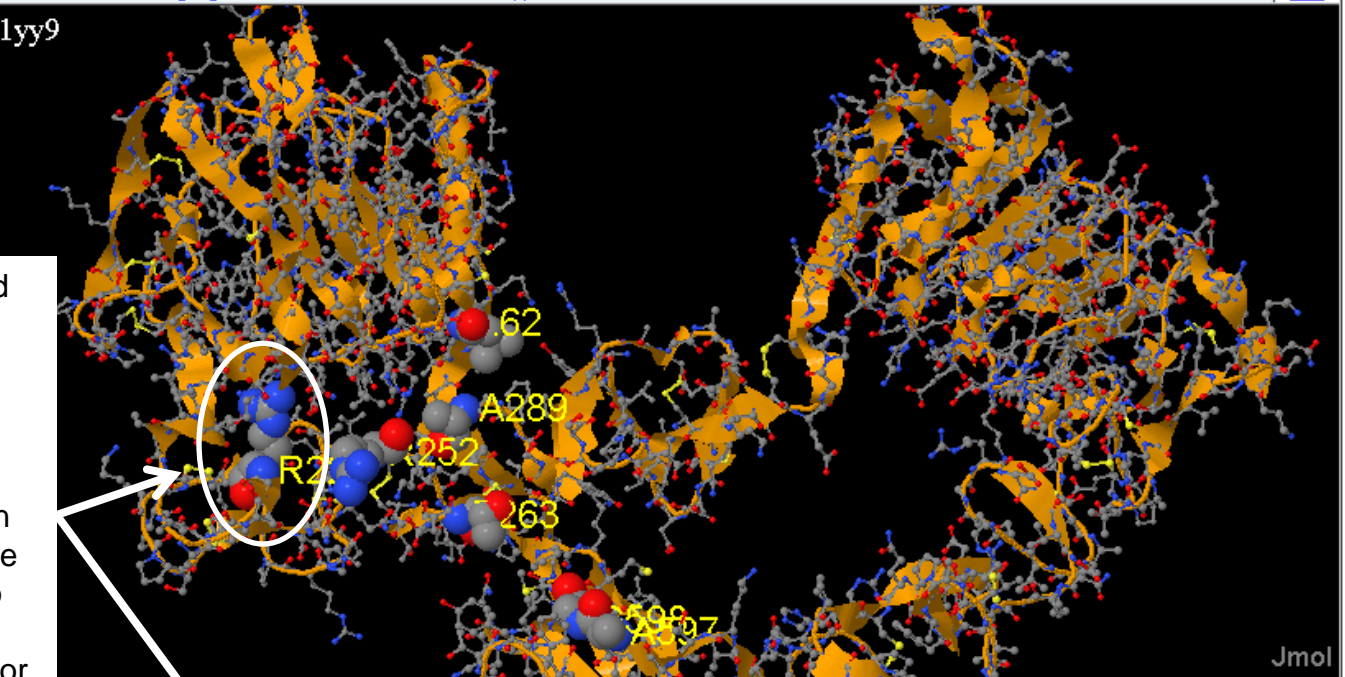
Protein Structure 3D Viewer - Mozilla Firefox

https://cgwb.nci.nih.gov/cgi-bin/3dviewer/ViewAA.cgi?proj=valid&gi=29725609&id=722&pdb=1yy9:A&sim=0.995&gstart=25&glen=613&phos=

National Cancer Institute U.S. National Institutes of Health | www.cancer.gov

GI: 29725609 Viewing region 25 - 638 99.5 % similar to Pdb: 1yy9 chain A 2 - 614 Help

1yy9



Big, highlighted atoms refer to the mutated amino acids (shown in red in the bottom panel). You can also click on the mutated amino acid (shown in red) to turn on or off a specific mutation

Sequence: Click on a letter once will turn on the spacefill. Double-click on the same letter will turn the spacefill off.

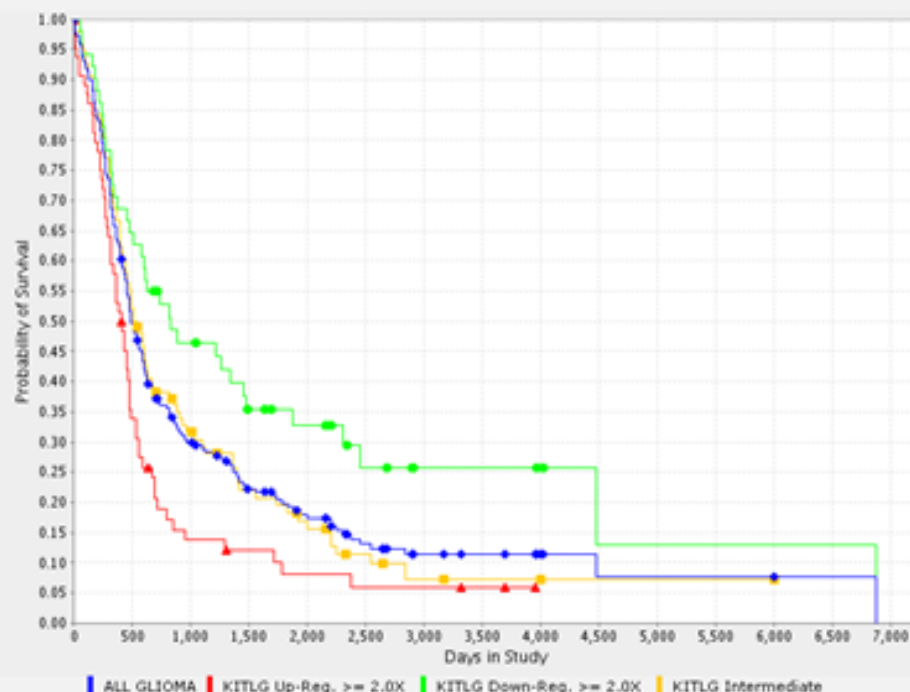
143	IVSSDFLSNMSMLFQNHGSCQKCDPSPNGSCWGAGEENCQKLTIIICAQQCSGR	213
214	AGCTGPRESDCLVLRKFRDEATCKDTCPPLMLYNPTTYQMDVNPPEGKYSFGAT	284
285	RACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFKNCTS	355
356	SFHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAFENLEIIRGRTKQHGGQ	426
427	RSLEISDGDVVISGNKNLCYANTINWKKLFGTSGQKTKIISNRGENKCKATGQV	497
498	DCVSCRNVSRGRECVDKCKLLEGEPREFVENSECIQCHPECLPQAMNITCTGRGP	568
569	KTCPAGVMGENNTLVWKYADAGHVCHLCHPNCTYGTCTGPGRLRGCT	614

Jmol

Jmol script completed cgwb.nci.nih.gov

# Gene expression analyses related to clinical outcome

**Kaplan-Meier Survival Plot for Samples with Differential KITLG Gene Expression**



[Upregulating Samples](#) [View Clinical Reports](#) [Downregulating samples](#) [Intermediate Samples](#)

## Statistical Report:

**KITLG**  
Reporter: 211124\_s\_at

### Number of samples in group:

Up-Regulated	64
Down-Regulated	51
Intermediate:	96

### Log-rank p-value(for significance of difference of survival between group of samples)

Up-Regulated vs. Intermediate:	0.0257
Up-Regulated vs. Down-Regulated	2.0E-4
Down-Regulated vs. Intermediate:	0.0046



## Administration:

- ◆ [View Results](#)
- ◆ [List Management](#)
- ◆ [Help](#)

## News:

- ◆ **Data Version**
- ◆ TCGA newsletter - March 2008
- ◆ Number of Patients - 110
- ◆ Number of Expression Arrays - 985
- ◆ Number of Copy Number Arrays - 361

## PatientDID Lists:

- ◆ ALL\_PATIENTS
- ◆ Low\_Survival
- ◆ Med\_Survival
- ◆ High\_Survival
- ◆ TP53\_SomaticMut...
- ◆ EGFR\_SomaticMut...
- ◆ PTEN\_SomaticMut...
- ◆ RB1\_SomaticMut...
- ◆ DST\_SomaticMut...
- ◆ NF1\_SomaticMut...
- ◆ CDKN2A\_Somatic...
- ◆ PIK3R1\_Somatic...
- ◆ CENPF\_SomaticM...
- ◆ ITGB3\_SomaticM...

## Gene Lists:

- ◆ TCGA Target Se...

## Reporter Lists:

# High order analyses: Pathways

## Pathways and Associated Anomalies

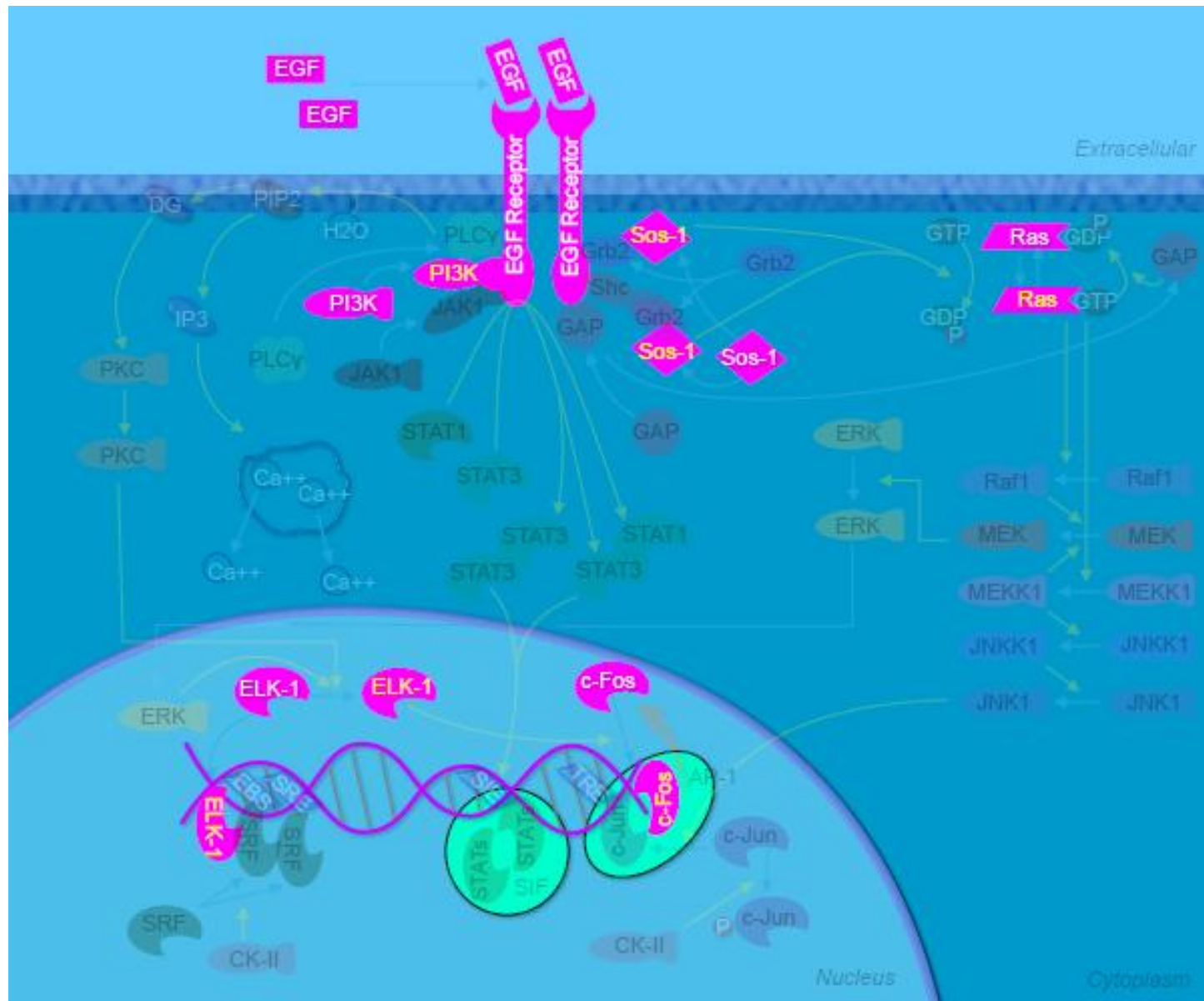
100 Rows Displayed

309 results found, displaying 1 to 100

Pathway	Any Anomaly/Agent	Mutated	Amplified	Deleted	Agents
<a href="#">ADP-Ribosylation Factor</a>					
<a href="#">AKAP95 Role in Mitosis and Chromosome Dynamics</a>					
<a href="#">AKT Signaling Pathway</a>					
<a href="#">ALK in Cardiac Myocytes</a>					
<a href="#">ATM Signaling Pathway</a>					
<a href="#">Acetylation and Deacetylation of RelA in Nucleus</a>					
<a href="#">Actions of Nitric Oxide in the Heart</a>					
<a href="#">Activation of CSK Inhibits Signaling through the T Cell Receptor</a>					
<a href="#">Activation of PKC through G-Protein Coupled Receptors</a>					
<a href="#">Activation of cAMP-dependent Protein Kinase, PKA</a>					
<a href="#">Acute Myocardial Infarction</a>					
<a href="#">Adhesion Molecules on Lymphocyte</a>					
<a href="#">Adhesion and Diapedesis of Granulocytes</a>					
<a href="#">Adhesion and Diapedesis of Lymphocytes</a>					
<a href="#">Aqirin in Postsynaptic Differentiation</a>					
<a href="#">AhR Signal Transduction Pathway</a>					
<a href="#">Alpha-synuclein and Parkin-mediated Proteolysis in Parkinson's Disease</a>					
<a href="#">Alternative Complement Pathway</a>					
<a href="#">Angiotensin II Mediated Activation of JNK Pathway via Pyk2 Dependent Signaling</a>					



# EGFR network mutation profile through CMA



# caBIG®: Power of Connection

